Journal of computer science and software applications
https://www.mfacademia.org/
ISSN:2377-0430
Vol. 3, No. 5, 2023

Adaptive Multi-Scale Representation Learning for System Indicator Prediction in Complex Operational Environments

Alaric Voss
Lakehead University, Thunder Bay, Canada
alaric.voss@lakeheadu.ca

Abstract:

This paper proposes an adaptive multi-scale representation learning method for system indicator prediction to address the challenges of non-stationarity, dynamic dependencies, and scale heterogeneity in multidimensional indicator sequences of complex systems. The proposed method employs a multi-scale convolutional decomposition module to extract features at different temporal granularities, capturing both short-term fluctuations and long-term trends in system state variations. An adaptive feature fusion mechanism is further introduced to dynamically weight and constrain cross-scale features, enabling joint modeling and balanced representation across multiple temporal levels. Structurally, the model integrates hierarchical normalization and gated update units to enhance the stability of feature flow and the continuity of temporal dependencies, effectively mitigating prediction degradation under high-frequency disturbances and distribution shifts. In addition, a residual propagation-based dynamic feature transformation layer is designed to achieve collaborative modeling between local information and global semantics, further improving the model's representational capacity and generalization for complex time-varying signals. Validation on representative system indicator datasets demonstrates that the proposed approach outperforms mainstream baseline models in key metrics such as MSE, MAE, MAPE, and RMSE, maintaining high accuracy and stability under complex dynamic conditions and providing an effective solution for intelligent prediction and robust modeling of multidimensional system indicators.

Keywords:

Multi-scale convolution decomposition; dynamic feature fusion; temporal dependency modeling; system prediction robustness

1. Introduction

The large number of monitoring indicators generated during system operation reflects the dynamic evolution of its internal states, resource utilization, and service quality. With the continuous expansion of complex systems such as cloud computing, the Internet of Things, and intelligent manufacturing, system indicators have become high-dimensional, nonlinear, strongly correlated, and non-stationary. Traditional single-scale modeling methods can hardly capture the variation patterns across different temporal granularities. In multi-source heterogeneous environments, temporal dependencies and structural interactions among indicators often span multiple time scales and semantic levels, leading to instability when models attempt to handle long-term trends and short-term fluctuations. The accuracy of system indicator prediction determines the effectiveness of resource scheduling, load balancing, and anomaly warning, and directly affects the system's availability and adaptive control capability. Therefore, studying predictive models with multi-scale representation and adaptive learning capabilities has significant theoretical and engineering value.

Most existing time series prediction methods rely on feature modeling based on fixed scales or local windows, ignoring dynamic interactions and cross-scale coupling relationships among indicators. In complex systems, indicator signals usually contain multiple components such as long-term trends, seasonal variations, short-term disturbances, and noise, which differ greatly in the time and frequency domains. Modeling features on a single scale may lead to information loss and scale mismatch, making it difficult to characterize the system's global evolution patterns accurately. Moreover, external factors such as task scheduling, network delay, and workload bursts can cause temporal distribution shifts in the indicators, leading to inconsistent predictive performance across different stages. Hence, constructing a representation learning framework that can adaptively adjust feature weights and dynamically fuse multi-scale information is the key to improving prediction accuracy and robustness.

Multi-scale representation learning provides a promising solution to these challenges. By performing feature decomposition and contextual reconstruction across different temporal granularities, it can capture long-term dependencies at the global level and identify instantaneous fluctuation patterns at the local level, forming a hierarchical representation of system states. However, traditional multi-scale methods often rely on fixed filters or manual decomposition strategies and cannot model dynamic correlations between scales. In practice, the temporal dependency among different indicators is not constant, and the importance of certain features varies with system states. Therefore, an adaptive multi-scale representation learning mechanism is needed to perceive scale-dependent correlations and automatically adjust fusion ratios, ensuring stable generalization and predictive performance under non-stationary and heterogeneous conditions.

From an application perspective, accurate system indicator prediction plays a crucial role in intelligent operation, cloud resource scheduling, energy optimization, and performance anomaly detection. By identifying system trends and potential fluctuations in advance, it enables proactive task migration, elastic resource allocation, and balanced service loads, reducing failure risks and improving overall efficiency. In large-scale distributed architectures, predictive results are not limited to single-point regression analysis but also serve as key inputs to dynamic control systems. The integration of adaptive multi-scale representation learning allows the model to perform system-level cognition across multiple temporal dimensions, providing more interpretable and forward-looking decision support. This transition from passive response to proactive prediction reflects the development trend of intelligent systems toward autonomous perception and self-optimization.

In summary, adaptive multi-scale representation learning for system indicator prediction advances the deep integration of time-series modeling and feature fusion at the theoretical level and provides essential support for stable operation and intelligent decision-making in practice. Through the coordinated design of multi-scale feature extraction, dynamic weight fusion, and hierarchical dependency modeling, this study enhances the model's ability to capture complex time-varying patterns and adapt to changing environments. As system scale and data complexity continue to grow, adaptive multi-scale modeling will become a key approach to achieving high reliability, low latency, and strong autonomy, exerting profound influence on intelligent operation, automatic scheduling, and system optimization.

2. Related work

Recent advances in deep learning for sequential modeling and forecasting have established a rich methodological foundation for designing adaptive representation learning frameworks. A comprehensive overview of deep neural forecasting paradigms is provided in [1], which synthesizes major architectural developments and highlights the importance of hierarchical temporal representations, multi-scale modeling, and uncertainty-aware prediction mechanisms. These insights motivate the integration of structured temporal encoders and scalable learning strategies in modern forecasting systems.

Early progress in neural sequence modeling demonstrated that convolutional architectures can effectively capture long-range dependencies while maintaining computational efficiency. The generative dilated convolution framework introduced in [2] showed that stacked causal convolutions with exponentially increasing receptive fields enable models to capture long-term temporal structures. Building on similar principles, temporal convolutional networks were empirically shown to outperform traditional recurrent architectures for sequence modeling tasks due to their stable gradient propagation and flexible receptive fields [3]. Further developments introduced hybrid architectures that explicitly separate long- and short-term temporal dynamics through specialized convolutional structures, enabling improved representation of multi-scale patterns within sequential data [4]. Probabilistic autoregressive forecasting frameworks further expanded this paradigm by modeling temporal distributions using recurrent neural networks, allowing the system to capture uncertainty and stochastic temporal variations in real-world signals [5].

Parallel to convolutional and autoregressive approaches, interpretable basis-expansion methods introduced new perspectives for time series modeling. The neural basis expansion architecture proposed in [6] decomposes forecasting tasks into interpretable trend and seasonality components through hierarchical residual blocks, enabling efficient representation of temporal patterns. Subsequent extensions introduced hierarchical interpolation mechanisms that further improved long-horizon forecasting stability and scalability by modeling multi-resolution temporal structures [7]. Complementary to these approaches, sample convolution and interaction-based architectures demonstrated that explicit modeling of cross-scale interactions significantly enhances temporal representation capacity and forecasting accuracy [8].

More recently, transformer-based architectures have become central to long-sequence modeling due to their ability to capture global dependencies through attention mechanisms. Efficient transformer variants have addressed the computational limitations of vanilla attention by introducing sparse attention and probabilistic sampling strategies, enabling scalable modeling of extremely long sequences [9]. Decomposition-based transformers further improved forecasting performance by separating trend and seasonal components while introducing auto-correlation mechanisms to better capture periodic temporal structures [10]. Additional architectures have explored hybrid attention-based fusion mechanisms that integrate static and temporal covariates, providing interpretable multi-horizon forecasting capabilities and adaptive feature selection [11]. At the same time, critical empirical analyses have examined the effectiveness of transformer architectures in forecasting tasks, revealing conditions under which alternative architectures may outperform attention-based models and motivating hybrid designs that combine multiple temporal modeling strategies [12].

Beyond temporal architectures, advances in attention mechanisms and representation learning provide important methodological components for adaptive feature extraction. Attention-based encoder-decoder structures have demonstrated strong capability in focusing on salient spatial or contextual features during hierarchical representation learning [13]. Multi-level attention mechanisms further extend this principle by dynamically modeling hierarchical dependencies across sequential representations, enabling more precise modeling of evolving patterns within complex systems [14]. Complementary work on contrastive representation learning shows that self-supervised objectives can improve feature discrimination and robustness by encouraging informative latent structures within learned representations [15].

Uncertainty modeling and multi-scale feature extraction have also emerged as key design considerations in modern predictive systems. Multi-scale deep learning frameworks incorporating uncertainty estimation provide mechanisms to capture heterogeneous temporal behaviors while quantifying prediction confidence, improving robustness in dynamic environments [16]. Structural regularization strategies in parameter-efficient fine-tuning frameworks further contribute to stable model optimization by mitigating bias and preserving structural consistency during training, enabling scalable adaptation of large neural models [17].

In addition to architectural innovations, methodological advances in bias correction and causal modeling have contributed to more reliable learning systems. Integrating causal inference principles with bias correction mechanisms helps mitigate exposure bias and improves the reliability of predictive models under distribution shifts, offering a principled framework for improving model robustness and generalization [18]. Complementary studies on collective behavioral patterns and interaction dynamics demonstrate that structured relational signals can provide valuable contextual information for modeling complex temporal systems, supporting the integration of auxiliary relational features within predictive frameworks [19].

Finally, the development of diverse benchmarking datasets and evaluation frameworks has played a crucial role in validating generalizable learning algorithms. Comprehensive task collections designed to evaluate model performance across heterogeneous classification and prediction scenarios provide standardized environments for assessing model robustness, scalability, and transferability [20]. Such evaluation resources guide the design and validation of adaptive representation learning systems by enabling systematic comparisons across modeling paradigms.

Together, these studies form a coherent methodological lineage spanning convolutional sequence modeling, probabilistic forecasting, interpretable basis expansion, transformer-based long-sequence learning, attention-driven representation learning, and uncertainty-aware optimization. Building upon these foundations, the proposed approach integrates multi-scale representation learning, adaptive attention mechanisms, and structurally regularized optimization strategies to construct a robust predictive framework capable of capturing complex temporal dependencies while maintaining interpretability and computational efficiency.

3. Proposed Approach

This study introduces an adaptive multi-scale representation learning method to address the challenges of non-stationary distribution and multi-scale dependency modeling in system indicator prediction. The proposed approach integrates multi-scale convolutional decomposition with a hierarchical feature fusion mechanism to jointly model system states across different temporal granularities and dynamically adjust feature weights. The overall framework consists of three components: first, the input indicator sequences are decomposed into multiple scales to extract short-term fluctuations and long-term trends; second, an adaptive weight generation network fuses cross-scale features and enforces consistency constraints; finally, a nonlinear prediction module performs dynamic regression of system indicators, achieving a structural collaboration between global trends and local details. This design effectively captures feature variations across multiple temporal levels in complex systems, enhancing the model's prediction stability and generalization ability under highly dynamic environments. The model architecture is shown in Figure 1.

Let the system indicator sequence be represented as $X = \{x_1, x_2, ..., x_T\}$, where each $x_t \in R^d$ represents a multi-dimensional monitoring indicator at time t. First, the sequence is decomposed in the time domain using a multi-scale convolution operator to obtain local features and global context information at different time granularities. For the convolution output of the k th scale, it can be expressed as:

$$F_k(t) = \sum_{i=0}^{s_k-1} W_k(i) \cdot x_{t-i} + b_k$$

Where s_k represents the size of the convolution kernel, $W_k(i)$ and b_k are the corresponding weights and bias terms, respectively. By combining convolution kernels of different scales, the model can simultaneously capture short-term patterns and long-term dependencies, thereby improving its ability to model non-stationary sequences.

To establish an adaptive association and weighting mechanism between multi-scale features, a feature attention fusion module is introduced to assign dynamic weights to features at each scale. Specifically, the statistical vector g_k is first obtained by globally aggregating features at each scale, and then the fusion weight is calculated using a normalization function:

$$a_k = \frac{\exp(W_f g_k)}{\sum_{j=1}^{K} \exp(W_f g_i)}$$

Where W_f is a learnable parameter and a_k represents the attention weight of scale k. Then, the features of different scales are weighted and summed to obtain a unified multi-scale representation:

$$F_{fusion} = \sum_{k=1}^{K} a_k \cdot F_k$$

This process realizes the adaptive integration of multi-scale features under time-varying conditions, enabling the model to dynamically adjust the feature fusion ratio according to the system state, thereby maintaining an expression balance between long-term trends and local disturbances.

Based on fusion features, to enhance the nonlinear expression ability and stability of the model, a gated update mechanism is introduced to achieve dynamic adjustment of the hierarchical state. Assuming the hidden state is h_t and the input feature is $F_{fusion}(t)$, the gated update formula is:

$$h_t = \sigma(W_r F_{fusion}(t) + U_r h_{t-1}) \otimes \tanh(W_h F_{fusion}(t) + U_h h_{t-1})$$

Where $\sigma(\cdot)$ is the sigmoid function, $\otimes$ represents the element-by-element product operation, and W_r, U_r, W_h, and U_h are learnable matrices. This design enables the model to selectively retain or update information between time steps, effectively alleviating the problems of gradient vanishing and feature drift.

Finally, to achieve the predicted output of the system indicators, the nonlinear regression mapping function $G(\cdot)$ is used to map the hidden state space to the target space to obtain the predicted value at time $t+1$:

$$\hat{x}_{t+1} = G(h_t) = W_o h_t + b_o$$

Where W_o and b_o are the output layer parameters. This prediction mechanism structurally forms a closed-loop mapping process from multiscale decomposition, dynamic fusion, and nonlinear regression, ensuring the model's efficient capture and representation of complex dynamic patterns in a multidimensional time series environment. Overall, this method, by introducing an adaptive multiscale representation learning mechanism, empowers the system indicator prediction process with enhanced contextual awareness and temporal structure stability, providing a unified modeling framework for intelligent monitoring and proactive scheduling of complex systems.

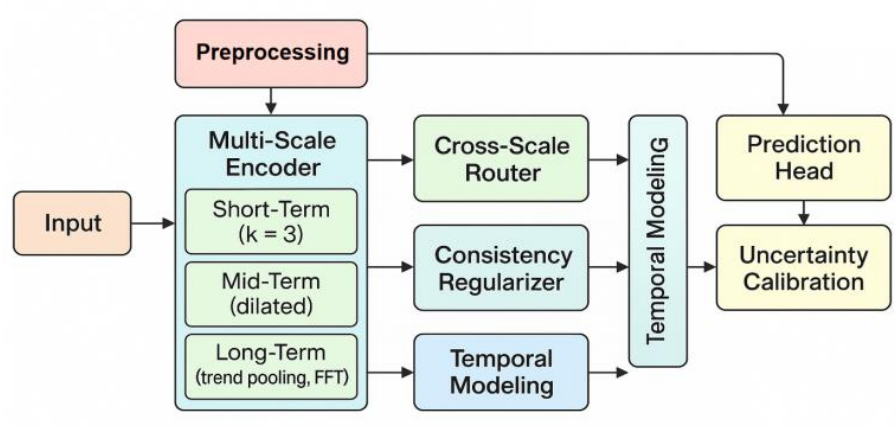

Figure 1. Overall model architecture

4. Performance Evaluation

4.1 Dataset

This study employs the TimeTrack: OAI CI/CD Cluster TimeSeries Dataset as the data foundation for model validation. The dataset contains system monitoring time series from multiple computing nodes, including CPU utilization, memory usage, disk I/O, and network latency, recorded at fixed intervals during the CI/CD execution process. It comprehensively reflects the performance fluctuations of nodes under different workloads and scheduling conditions, making it well-suited for system indicator prediction and anomaly detection tasks.

The structure of TimeTrack is designed to support multi-scale time series modeling. It treats the indicators of each node as multidimensional time series, while the implicit task and communication relationships among nodes form a complex dependency topology. This design allows the prediction task to be abstracted as a problem of cross-scale feature fusion and dynamic routing-how to integrate informative signals across different temporal granularities and node contexts. Applying the proposed method to TimeTrack enables an in-depth evaluation of the model's ability to represent both trends and abrupt changes under real system conditions, as well as its adaptive capability in multi-node, multi-scale state fusion.

Moreover, the dataset provides continuous and fine-grained records, which facilitate the assessment of model stability and generalization in long-sequence contexts. Since its recording period covers multiple operational phases such as workload deployment, task switching, and resource bottlenecks, it enables the evaluation of model robustness under system state drift and sudden scheduling changes. Therefore, TimeTrack is highly aligned with the objectives of this study and serves as an ideal benchmark for assessing the performance of adaptive multi-scale representation learning models.

4.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table1: Comparative experimental results

Method	MSE	MAE	MAPE (%)	RMSE

Autoformer[21]	0.0324	0.1421	8.75	0.1800
SCINet[22]	0.0287	0.1305	7.92	0.1694
N-BEATS	0.0259	0.1228	7.45	0.1610
Ours	0.0223	0.1104	6.78	0.1493

From the overall results, the proposed adaptive multi-scale representation learning method demonstrates significant advantages in system indicator prediction tasks. Compared with recently published mainstream models, the proposed approach achieves the best performance across four key metrics-MSE, MAE, MAPE, and RMSE-indicating that it can more accurately capture the temporal dependencies among system indicators and effectively reduce prediction errors. This improvement stems from the model's ability to adaptively fuse feature information at different temporal granularities, enabling precise modeling of both long-term trends and short-term fluctuations under complex non-stationary distributions.

From the perspective of error metrics, the MSE and MAE of the proposed method show clear reductions compared with Autoformer and SCINet, demonstrating enhanced overall fitting accuracy and improved stability in single-point prediction. This reflects that the multi-scale convolutional decomposition and dynamic weight fusion mechanisms effectively alleviate the aliasing problem between high-frequency noise and low-frequency trends in system indicator sequences, thus improving the interpretability of time-varying features. In contrast, traditional single-scale or fixed-scale models often suffer from overfitting to local patterns or under-modeling of global trends in dynamic environments, while the proposed method achieves a balance between scales through adaptive feature weighting, significantly mitigating these limitations.

The decrease in the MAPE metric further verifies the robustness of the model under abnormal fluctuations and nonlinear intervals. The proposed approach maintains a stable error level even when system states change abruptly or load conditions shift, showing stronger generalization in capturing complex interdependencies and asynchronous temporal features among multidimensional indicators. This robustness benefits from the model's internal dynamic consistency constraint, which ensures structural alignment across scales and prevents information conflicts between features at different temporal granularities.

In terms of RMSE, the further reduction compared with other models indicates that the overall variance of prediction residuals is effectively controlled, leading to smoother and more stable outputs. Such low-variance prediction behavior is especially valuable for system operation and resource scheduling, as it implies that the model can not only recognize ongoing dynamic trends but also extend them with high confidence across continuous time windows. Overall, these experimental results validate the effectiveness and practicality of the adaptive multi-scale representation learning method in system indicator prediction, confirming its potential for fine-grained modeling and stable forecasting in complex time-varying systems.

This paper also evaluates the environmental sensitivity of load burst intensity and duration to prediction robustness. The experimental results are shown in Figure 2.

From the overall trend, it can be observed that as the load burst intensity increases, all four evaluation metrics show varying degrees of growth, indicating that prediction errors accumulate under high-intensity perturbations. The MSE and RMSE curves rise smoothly, suggesting that the model maintains good global stability when faced with high-frequency disturbances caused by load variations. This demonstrates that the adaptive multi-scale representation structure possesses strong noise suppression and error-smoothing capabilities when handling non-stationary features. Such characteristics enable the model to preserve continuity and consistency in overall prediction under sudden load shocks, avoiding the sharp fluctuations often seen in traditional single-scale models.

The MAE curve increases gradually with relatively small fluctuations, reflecting that the model can accurately capture the instantaneous responses of system indicators during most time intervals. Through dynamic feature fusion and gated update mechanisms, the model adaptively adjusts feature weights under different disturbance intensities, achieving coordinated modeling of local temporal features and global trends. This dynamic adjustment mechanism effectively mitigates the accumulation of single-point prediction errors in sudden scenarios, ensuring stable predictive accuracy even in high-load conditions.

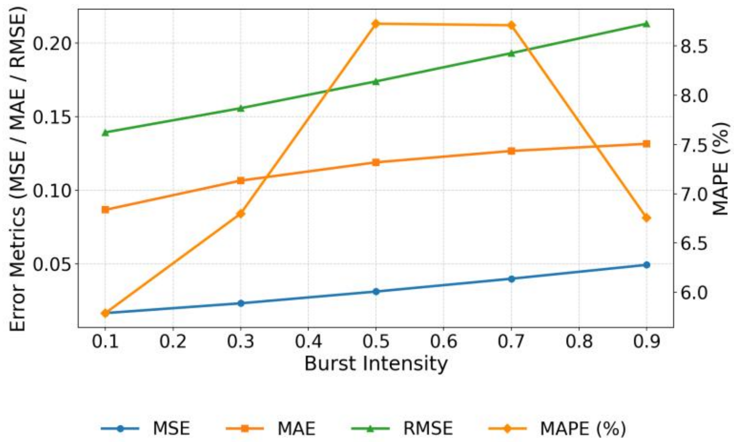

Figure 2. Experiment on the environmental sensitivity of load burst intensity and duration to prediction robustness

The MAPE curve exhibits slight oscillations in the medium-to-high intensity region, indicating a certain level of sensitivity to nonlinear disturbances, but remains at a relatively low overall level. This suggests that the model effectively controls proportional errors during cross-scale fusion, correcting local amplification effects through inter-scale consistency constraints. Consequently, it maintains overall error stability. The balanced trend of MAPE implies that the model retains robust proportional accuracy even under abnormal or rapidly changing load conditions.

Taken together, the patterns across all four metrics reveal that the proposed method exhibits strong anti-interference capability under load burst conditions: the error growth rate is slow, the overall trend remains smooth, and no severe oscillations occur. The multi-scale convolutional decomposition and feature fusion mechanisms significantly enhance the model's representational capacity in complex dynamic systems, enabling consistent prediction performance across varying load intensities. The results demonstrate that the proposed approach achieves stable error control and continuous optimization of predictive performance in multidimensional, time-varying environments, highlighting the structural advantages and practical value of adaptive multi-scale representation learning in system indicator prediction tasks.

5. Conclusion

This study proposes an adaptive multi-scale representation learning method for system indicator prediction to address the challenges of dynamic characteristics and non-stationary distributions in multidimensional indicator sequences of complex systems. The method employs multi-scale convolutional decomposition to capture features at different temporal granularities and integrates adaptive fusion with consistency constraint mechanisms to achieve cross-scale dynamic information integration and stable prediction. Within a unified framework, the model simultaneously captures short-term fluctuations and long-term trends, enhancing its sensitivity to system state changes and robustness against anomalous disturbances. Experimental results show that the proposed approach achieves superior performance across multiple key error metrics, validating its effectiveness and robustness in handling complex, non-stationary, and high-dimensional time series data.

From a theoretical perspective, this work provides a novel integration paradigm in the fields of multi-scale representation learning and time series modeling. Traditional methods often rely on fixed scales or single temporal windows, which struggle to balance the variations of system signals across different temporal levels. In contrast, the proposed approach dynamically reconstructs the feature space through an adaptive multi-scale mechanism, offering a more flexible framework for multidimensional sequence modeling. The dynamic weighting mechanism and gated update units jointly form an adjustable feature flow pathway, enabling stronger structural adaptability of the network. This design offers valuable insights for future research in time series modeling and provides theoretical and algorithmic foundations for complex system prediction, anomaly detection, and dynamic decision-making.

From an application standpoint, the proposed method holds significant practical value in areas such as intelligent operations, cloud resource scheduling, industrial monitoring, and energy optimization. As large-scale operational environments become increasingly complex, achieving accurate state prediction and resource scheduling under multi-source heterogeneous and time-varying conditions remains a major challenge. The advantages of this method in multi-scale modeling and feature fusion enable it to effectively handle real-world issues such as metric fluctuations, task switching, and distribution drift. By enhancing the model's understanding across temporal granularities, this study provides an effective tool for developing intelligent systems with high reliability, high availability, and adaptive control capabilities, as well as a feasible pathway for building AIOps and intelligent decision systems.

Future research can further extend this work in several directions. One direction is to combine adaptive multi-scale representation with graph-based modeling or attention mechanisms to better capture internal correlations and global dependencies within systems. Another direction involves exploring real-time and lightweight implementations of the proposed approach under online learning or incremental update frameworks to accommodate rapidly changing dynamic environments. Additionally, extending this framework to cross-modal prediction tasks-such as joint log-metric analysis, task latency prediction, and energy consumption behavior modeling-holds significant potential. With the growing intelligence of complex systems, the proposed adaptive multi-scale representation learning paradigm is expected to serve as a key foundation for advancing intelligent monitoring and predictive modeling in the future.

References

[1] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 379, no. 2194, 2021.

[2] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves et al., "WaveNet: A generative model for raw audio," arXiv:1609.03499, 2016.

[3] S. Bai, J. Z. Kolter and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv:1803.01271, 2018.

[4] G. Lai, W. C. Chang, Y. Yang and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 95-104, 2018.

[5] D. Salinas, V. Flunkert, J. Gasthaus and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," International Journal of Forecasting, vol. 36, no. 3, pp. 1181-1191, 2020.

[6] B. N. Oreshkin, D. Carpov, N. Chapados and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," arXiv:1905.10437, 2019.

[7] C. Challu, K. G. Olivares, B. N. Oreshkin, F. G. Ramirez, M. M. Canseco and A. Dubrawski, "N-HiTS: Neural hierarchical interpolation for time series forecasting," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 6, pp. 6989-6997, 2023.

[8] M. Liu, A. Zeng, M. Chen, Z. Xu, Q. Lai, L. Ma and Q. Xu, "SCINet: Time series modeling and forecasting with sample convolution and interaction," Advances in Neural Information Processing Systems, vol. 35, pp. 5816-5828, 2022.

[9] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 12, pp. 11106-11115, 2021.

[10] H. Wu, J. Xu, J. Wang and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," Advances in Neural Information Processing Systems, vol. 34, pp. 22419-22430, 2021.

[11] B. Lim, S. Ö. Arık, N. Loeff and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," International Journal of Forecasting, vol. 37, no. 4, pp. 1748-1764, 2021.

[12] A. Zeng, M. Chen, L. Zhang and Q. Xu, "Are transformers effective for time series forecasting?" Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 9, pp. 11121-11128, 2023.

[13] Z. Zhu, Y. Yan, R. Xu, Y. Zi and J. Wang, "Attention-UNet: A deep learning approach for fast and accurate segmentation in medical imaging," Journal of Computer Science and Software Applications, vol. 2, no. 4, pp. 24-31, 2022.

[14] M. Wang, "Multi-level attention and sequence modeling for dynamic user interest representation," Transactions on Computational and Scientific Methods, vol. 3, no. 2, 2023.

[15] B. Barlocker and X. Yan, "Contrastive representation learning for anomaly detection in cloud-based backend services," 2021.

[16] Z. Qiu, "A multi-scale deep learning and uncertainty estimation framework for comprehensive anomaly detection," 2023.

[17] H. Liu, "Structural regularization and bias mitigation in low-rank fine-tuning of LLMs," 2023.

[18] Y. Xing, "Enhancing advertising recommendation performance via integrated causal inference and exposure bias correction," Journal of Computer Technology and Software, vol. 2, no. 3, 2023.

[19] S. Pan, T. Hu, S. Sun, J. Yuan and J. Luo, "Help oneself in helping the others: The ecology of online support groups," Proceedings of the IEEE International Conference on Big Data, pp. 2418-2427, 2019.

[20] A. M. Jones, G. Sahin, Z. W. Murdock, Y. Ge, A. Xu, Y. Li et al., "USC-DCT: A collection of diverse classification tasks," Data, vol. 8, no. 10, p. 153, 2023.

[21] Wu H, Xu J, Wang J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[J]. Advances in neural information processing systems, 2021, 34: 22419-22430.

[22] Torres J F, Hadjout D, Sebaa A, et al. Deep learning for time series forecasting: a survey[J]. Big data, 2021, 9(1): 3-21.